DOCUMENT RESUME

ED 088 749 SO 007 098

AUTHOR Taylor, Bob L.
TITLE The National Assessment Model.
PUB DATE [74]
NOTE 14p.; Paper presented at the National Council for Social Studies, General Session of College and University Faculty (San Francisco, November, 1973)

EDRS PRICE MF-$0.75 HC-$1.50
DESCRIPTORS *Academic Achievement; Behavioral Objectives; *Citizenship; *Data Collection; *Educational Assessment; Educational Research; Educational Testing; Evaluation; Evaluation Methods; *National Surveys; Program Descriptions; Program Effectiveness; Research Methodology; Research Utilization; Sampling; School Surveys; Writing Exercises
IDENTIFIERS *National Assessment

ABSTRACT

The National Assessment is a plan for a systematic survey of knowledges, skills, understandings, and attitudes. The development of citizenship assessment described in this paper typifies the processes undergone for each subject--art, career and occupational development, citizenship, literature, mathematics, music, reading, science, social studies, and writing. Of the following developmental components, emphasis has been placed on the first five, although the last two are now receiving greater attention: 1) development of objectives with the participation of educators and laymen; 2) development of exercises based on the objectives and representing various techniques and levels of difficulty; 3) determination of a probability population sample representing region, size of community, and sociceconomic status; 4) administration of the exercises to the sample of 9-, 13-, and 17-year olds and an adult group aged 26-35; 5) scoring and analysis of the exercises by machine and trained evaluators; 6) reporting and dissemination; and 7) utilization. Components 6 and 7 describe the census-like reporting of data without interpretation. National Assessment produced data based on a broad range of objectives from higher cognitive levels to the affective domain and not simply on factual knowledge. (JH)

ED 088749

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

THE NATIONAL ASSESSMENT MODEL

by

Bob L. Taylor
University of Colorado

Under National Assessment, several volumes of data have already been collected and published. There no longer is a debate concerning the pros and cons of having a National Assessment program, for the results of the early assessments are now available to the public. Educators need to study and interpret the results available as a base for decision making in education. Along with this, they need to understand how these data were collected and what were the problems and limitations of the data gathering process.

National Assessment is a plan for the systematic, census-like survey of knowledges, skills, understandings, and attitudes. It is an information gathering plan aimed at providing both educators and the lay public with information concerning the level of achievement in selected subject areas for students and young adults. The goal is to provide information that will be used to improve education. It is concerned with the achievement status of four age levels in ten different subject areas. The subject areas selected for assessment were: Art, Career and Occupational Development, Citizenship, Literature, Mathematics, Music, Reading, Science, Social Studies, and Writing. The basic task of this paper is to describe the model used by National Assessment for data gathering and reporting on the Citizenship area. It is not within the scope

SP 007 090

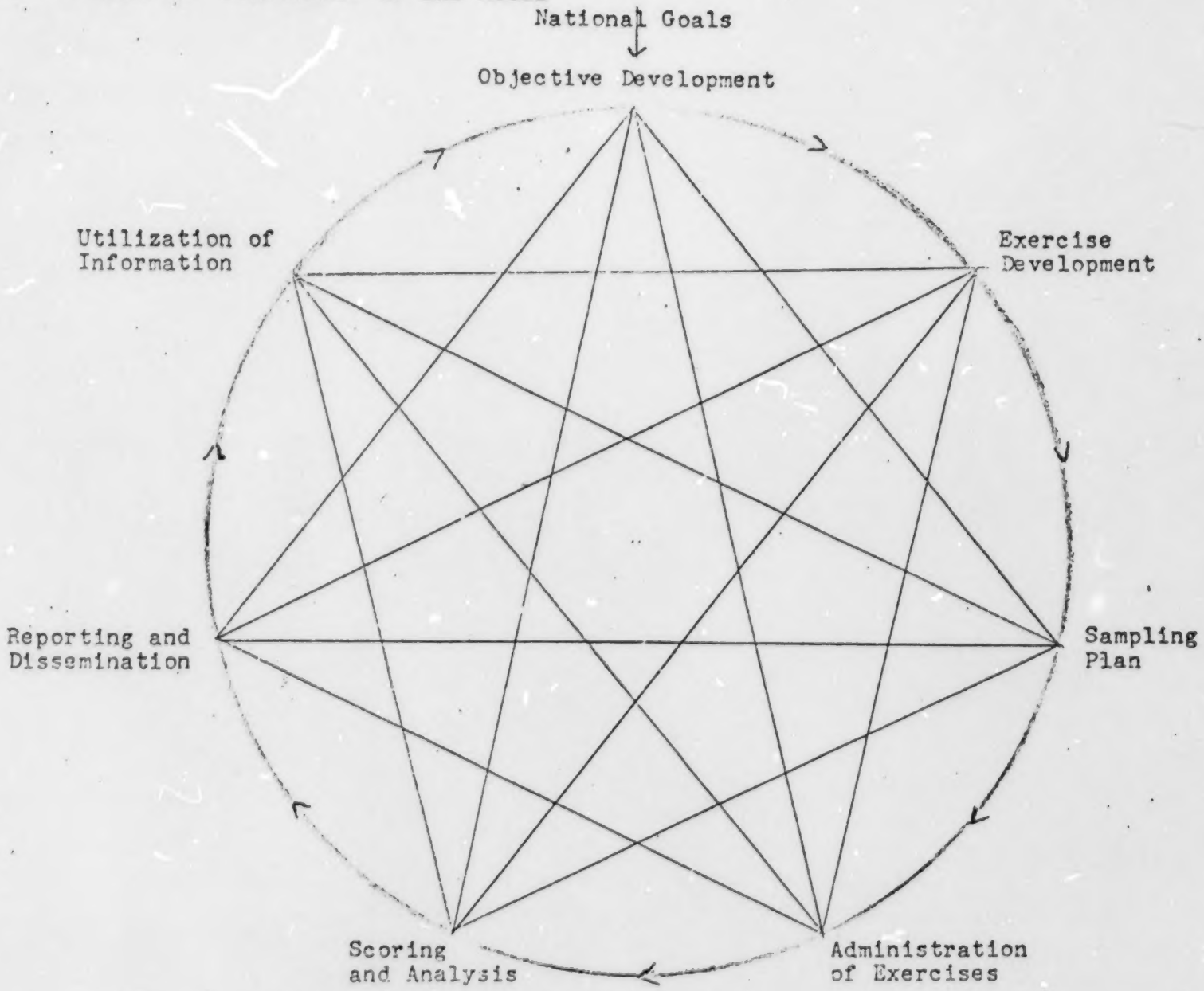
of this paper to criticize the model with respect to technical flaws; hence, the model is described and discussed without reference to any potential problems of design.

THE MODEL

The model is in the continuous process of being refined and improved, thus only the basic components of the model are presented in Diagram I. A circular scheme is used in presenting it since, in reality, its actual application may be initiated with any one of the components. Also, in its actual application, there are continual interactions between and among the various components. While theoretically the process starts with the refinement of overall national goals into specific subject matter, behavioral objectives, and progresses in logical sequence through to the final Utilization of the Information, in practice there is much greater freedom with respect to the utilization of the components.

The model for the Citizenship Assessment is presented here in outline form with a fairly detailed description of its components. As presented in Diagram I, there are seven basic components identified in the model: Objectives Development, Exercises Development, Sampling Plan, Administration of Exercises, Scoring and Analysis, Reporting and Dissemination, and Utilization of Information. While many of the fine points of the model are not developed in the following outline, it is described in sufficient detail to give the reader a good understanding of how the data were collected and what implications might result from these data. The number of subtopics in the model and their distribution indicate that the major

DIAGRAM I. COMPONENTS OF THE MODEL



efforts of National Assessment have been with the first five components. The last three components have been areas of controversy and, therefore, have received less attention until recently.

Outline of the Assessment Model for Citizenship[1]

I. Objectives Development[2,3,4,5]

A. The task of developing objectives in the field of citizenship was awarded to the American Institute for Research of Palo Alto, California. These criteria were used in examining the objectives:

(a) They were considered important by scholars.
(b) They were accepted as an educational task by the school.
(c) They were considered desirable by thoughtful lay citizens.

Scholars reviewed the objectives for authenticity with respect to their subject fields; school people reviewed the objectives in terms of their actual emphasis in their schools; and laymen reviewed them in terms of their experiences with regard to their value in life.

B. The American Institute for Research staff reviewed previous lists of citizenship objectives and boiled these down to one comprehensive list of 20 objectives.

---

[1]Womer, Frank B., What Is National Assessment? National Assessment of Educational Progress, Denver, Colo., 1970.

[2]Norris, Eleanor L. (Ed.), Citizenship Objectives, Committee on Assessing the Progress of Education, Ann Arbor, Mich., 1969.

[3]Campbell, Vincent N., et al, Citizenship Objectives for 1974-75 Assessment, Education Commission of the States, National Assessment of Educational Progress, Denver, Colo., 1972.

[4]Campbell, Vincent N. and Daryl G. Nichols, "National Assessment of Citizenship Education," Social Education 32:279-81, June, 1969.

[5]Campbell, Vincent N., et al, Report 2, Citizenship: National Results, Education Commission of the States, National Assessment of Educational Progress, Denver, Colo., November, 1970.

C. Outstanding local teachers familiar with each target-age group (9, 13, 17, adult), working with the American Institute for Research staff, broke down each general objective into the most germane behaviors deemed appropriate as goals for a given age group.

D. A selected group of students and adults in each age group was asked by the American Institute for Research staff to recall and describe outstanding citizens of their acquaintance and specific incidents reflecting good and poor citizenship. These incidents and descriptions, about 1,000, were used to check the completeness of the initial list of objectives.

E. The objectives were stated on three levels (general objectives, sub-objectives, and behavioral age illustrations or statements). The results were summarized for each age group.

F. The revised list of objectives, broken down into important behaviors, was then worked over for three days by a panel of national leaders in citizenship education and related social sciences.

G. A group of persons in various roles from selected California communities reviewed the objectives and made suggestions. These included public and private school administrators, counselors, teachers, a judge, a county planner, labor and business leaders, and social scientists.

H. The objectives were then reviewed by panels of laymen. Eleven lay review panels representing four geographic areas of the country and three different community sizes were used. Each panel spent two days reviewing the objectives based on these two questions: "Is this something important for people to learn today?" and "Is this something I would like to have my children learn?"

II. Exercises Development[6,7,8]

---

[6]Womer, op. cit.

[7]Gadway, Charles J. (Ed.) Reading and Literature: General Information Yearbook, Education Commission of the States, Report 02-GIY, National Assessment of Educational Progress, Denver, Colo., May, 1972.

[8]Finley, Carmen J. and Frances S. Berdie, The National Assessment Approach to Exercise Development, National Assessment of Educational Progress, Ann Arbor, Mich., 1970.

A. The production of the exercises was initiated by the American Institute for Research in 1966. The exercises were developed to cover all of the major objectives and to represent the selected content areas. Many exercises required the use of interview techniques, as well as the usual pencil and paper exercises. Also, self report and group task exercises were used.

B. Because National Assessment intends to describe what people in an age group know, the exercises were written to reflect three difficulty levels--reporting knowledge or skills common to almost all persons in an age group, reporting skills or understandings of a typical member of an age group, and reporting understandings or knowledge developed by the most able persons in an age group.

C. All exercises were developed to meet these criteria: content validity, clarity, functional exercise format, clustering exercises based on a single set of stimulus materials, directionality of response, difficulty level, content sampling, and overlap between age groups. The exercises were direct measures of some pieces of knowledge, understandings, attitudes, or skills which were mentioned in one or more of the objectives.

D. The exercises were reviewed by panels of lay persons for clarity, meaningfulness, and invasion of privacy.

E. There was a tryout of the exercises involving representatives of groups in the actual assessment--regions, communities, races, sexes, and age groups. Following the tryouts, the American Institute for Research staff and subject matter specialists reviewed the tryout data and made needed revisions.

F. A committee of subject matter specialists, measurement specialists, and National Assessment staff members rated the exercises to be included in the packages according to a set of criteria, and based on the ratings the exercises were selected for use.

G. The selected exercises were reviewed by U. S. Office of Education personnel for any infringement of privacy on the part of the respondents or possible offensiveness.

H. Since there were about 160 minutes of testing time available for each age group in each subject area, the exercises used were only a small sample of the potential number of exercises. The exercises were assembled into administrative units (packages) for groups up to 12 persons.

III. Sampling Plan[9,10]

A. The sampling plan was subcontracted to Research Triangle Institute, Raleigh, North Carolina. A multi-stage design was used which was stratified by region, size of community, and socio-economic status. This was a probability sample which allowed researchers to collect data from a small sample of the population and to infer from that sample certain characteristics of the entire population.

B. The populations for assessment were all 9 year olds, all 13 year olds, all 17 year olds, and all young adults 26 through 35 years old in the 50 States plus the District of Columbia. The only exceptions were the exclusions of institutionalized individuals of these ages--those in hospitals, prisons, and others who could not be reached.

C. For ages 9 and 13, a school sample only was used and for the 26 through 35 age group a household sample only was used. For the 17 year olds, both a school and a household sample were used.

D. The entire country was divided into population areas as follows: cities, counties exclusive of cities, and pseudo-counties--two or more counties were put together when the population of a single county was less than 16,000. Each population unit of 16,000 residents was assigned a number.

E. The country also was divided into four geographic regions: Northeast, Southeast, Central, and West.

F. Each geographic region was divided into communities of four types: large cities of above 180,000 population, urban fringe, middle-sized cities between 25,000 to 180,000 population, and small town-rural of under 25,000 population.

G. The 52 sampling units for each geographic area were spread across the four community types in a fashion proportional to their population in relation to the area population.

---

[9]Norris, Eleanor L., et al, Report 1, 1969-1970 Science: National Results and Illustrations of Group Comparisons, J. R. Chrony and D. G. Horvitz, "Structure of Sampling and Weighting," Appendix C, Education Commission of the States, National Assessment of Educational Progress, Denver, Colo., July, 1970.

[10]Norris, Citizenship Objectives, op. cit.

H. To insure comparable representation from each part of the country, an equal number of sampling units was selected from each geographic region--52 from each of the four regions for a total of 208.

I. The actual sampling units chosen were selected at random. This plan did not guarantee that all 50 States would be included in the sample. This was not a survey objective, but later the design was changed so each state was included in the sample.

J. In each sampling unit selected, all school buildings enrolling students of the sample ages (public, private, and parochial) were identified.

K. The plan for schools was to select units of approximately 250 to 350 pupils for each age group and from at least two different buildings within each sampling unit for each age group.

L. Each cooperating building principal provided a list of names of students in the building from the specific age groups. This list was used for the final random selection of students to take the assessment exercises from that building.

M. Information about the areas was obtained from the U. S. census data. In order to report reliable information for lower socio-economic status groups, these groups were oversampled. There was a disproportionate number of schools from lower socio-economic status areas included. In the overall results, the data from the lower socio-economic areas were given the percentage value in which they occurred in the total population.

N. From each of the 208 geographical samples, 100 adults, ages 26 through 35, were randomly selected using the following procedures. Each of the 208 geographic samples was divided into equal secondary sampling units. Then ten secondary sampling units were randomly selected from the total 208 samples. Interviewers then personally contacted the people in the chosen secondary sampling units of the 26 through 35 age group and out-of-school 17 year olds. These persons were asked to participate in the assessment.

O. Individuals were classified as black, white, and other on the basis of information provided by the school or by observation. Results were given for black and white only. The number of individuals classified as other was too small to produce reliable results.

IV. Administration of Exercises[11,12]

---

[11] Womer, op. cit.

[12] Gadway, op. cit.

A. The administration of the exercises was subcontracted to Research Triangle Institute in the East and to Measurement Research Center of Westinghouse Learning Corporation, Iowa City, Iowa, in the West. Cooperation of schools was obtained by first contacting officials at the state and then at school district levels. There was above 90 per cent cooperation by schools.

Adults and out-of-school 17 year olds were contacted by a personal door-to-door household canvass. Each out-of-school participant was contacted individually. All had the right to refuse to cooperate.

B. A full-time trained staff of 27 district supervisors managed the field work. They were assigned to different geographical areas of the United States. They contacted schools and recruited and trained local teachers to help in the administration of the exercises in schools and recruited and trained other available persons for the out-of-school administration.

C. In the schools, students from a single age group from different classes were brought together in a room for exercise administration. Group size was at least 8 and usually 12 students.

D. The exercises were organized in packages which contained exercises from two or three different subject areas at a single age level. No one person took all the exercises in his age group. Age groups were assessed at different times of the year.

E. In packages administered to groups, taped directions and taped readings of the exercises were used in addition to printed packages. This was done to establish consistency in timing and administration plus to provide for nonreaders.

F. Several packages at ages 9, 13, and 17 consisted of exercises that were given by exercise administrators to one individual at a time. The administration of all the packages for the adult assessment was done by interviews.

G. Each package required about 50 minutes of administrative time. Each person took only one package with the exception of the out-of-school, 17 year olds who were asked to take four or five packages each since they were the most difficult and expensive group to locate.

H. Students' names were confidential and did not appear on any packages. The name roster was kept at the building level and used only in the organization of the in-school sampling.

V. Scoring and Analysis[13,14]

A. The scoring and analysis of the exercises were subcontracted to Measurement Research Center of Westinghouse Learning Corporation, Iowa City, Iowa.

B. The multiple-choice exercises were scored and recorded routinely by machine.

C. The openended exercises were scored by trained professionals using a key of acceptable and unacceptable achievements in terms of the objectives.

D. Results were reported for each goal. Also, the results were reported both as the percentage of any group of respondents making the desired responses to an exercise and as the difference between the percentage of a group making the desired responses and the corresponding national percentage.

E. In the assessment, there was a lack of proportionality among characteristics used in the comparison of groups, such as color, sex, parental education. A statistical procedure, balancing, was used to correct for this problem in the comparative analysis of the data. Balancing is a procedure to examine the performance of groups classified on one characteristic adjusting for the fact that these groups differ on a specified set of other characteristics.

VI. Reporting and Dissemination[15,16,17]

A. The reporting of results was directed to subject matter specialists, professional educators, and informed laymen. Multiple reports were developed to serve these different audiences.

B. Approximately 40 per cent of the exercises were reported at the end of each assessment year. Not all exercises were reported since they were to be used over again in future assessments in order to measure change by means of comparing the results on the uncontaminated exercises.

C. The exercises released for publication were selected to be representative of all exercises administered as well as the results received on the assessment.

---

[13]Womer, op. cit.

[14]Gadway, op. cit.

[15]Womer, op. cit.

[16]Campbell, Report 2, Citizenship: National Results, op. cit.

[17]Gadway, op. cit.

D. Reporting was done by 9, 13, 17, and 26 through 35 age groups. Since the same exercises were used with different age groups, there was comparable data across two or more age levels.

E. Reporting was also done by groups within the categories of regions, community types, sex, socio-economic status, and white, black, and other.

F. Final reports were printed with a short description of the exercises, the national percentage of success, and group differences from the national percentage of success for each exercise. This was done without any interpretation of results.

G. Both observed and balanced results for all exercises and by groups were reported. The effects of balancing on measured characteristics such as sex and region were included in the report.

H. There were no scores reported for individuals. No single individual took more than one twelfth of the exercises, and no individual took a package that sampled only a single subject area.

I. Results were reported through the media: written word, radio, television, films, and personal reports.

VII. Utilization of Information[18,19]

A. The results provided information for educational decision making. For example, considering the somewhat lower performance of the Southeast Region on the Citizenship results, school boards in that region might decide to put greater stress in their school programs on citizenship skills, understandings, and attitudes.[20]

---

[18]Womer, op. cit.

[19]Conway, Larry E., "Some Implications of the National Assessment Model and Data for State and Local Education," Paper Presented at the 1973 Annual Meeting of the AERA, New Orleans, Louisiana, February 26, 1973.

[20]Campbell, Vincent N., Manford J. Ferris, and Daryl G. Nichols, National Assessment Report 6, 1969-1970 Citizenship: Group Results for Sex, Region, and Size of Community, Education Commission of the States, National Assessment of Educational Progress, Denver, Colo., July, 1971.

B. The results raised many questions which may lead to other investigations. For example, in making comparisons of all Citizenship results combined, it was found that the Extreme Affluent Suburbs showed substantial median advantages at all ages and that the Extreme Rural and Extreme Inner City showed substantial deficits at all ages. Here are discrepancies in performance which need to have causal studies conducted on them from the perspectives of different disciplines such as political science, sociology, economics, and education.[21] A second example is that assessment could expand into other educational areas, such as assess a sample of 5 year olds for a basis of comparison, break results down by states, or expand into new subject areas.

C. The results of several cycles should provide evidence of the change in knowledge, skills, understandings, and attitudes in the age groups as they relate to educational objectives.

D. School administrators can make comparisons between groups, and may improve student performance from the information gained in this manner.

From this review of the model, it is evident that the National Assessment staff has put a great deal of effort and know-how into the design, plus the development of each of the components. Each of the components reflects the best available information in the field of assessment.

SUMMARY

A complicated but logical process has evolved for identifying the behavioral objectives for the specific subject areas. The use of laymen along with teachers and subject matter specialists added validity checks on these objectives. Considerable inventive talent was displayed in writing the exercises for assessing some of the

---

[21] Norris, Eleanor L., Vincent N. Campbell, Manford J. Ferris, and Carmen J. Finley, National Assessment Report 9, 1969-1970 Citizenship: Group Results for Parental Education, Color, Size, and Type of Community, Education Commission of the States, National Assessment of Educational Progress, Denver, Colo., May, 1972.

noncognitive type of objectives. The use of the tape-paced administration was a valuable innovation in the field of testing. In this procedure, a complete set of directions for the administration of a package of exercises is recorded on an audio tape. By playing the tape, the directions are always given in the same manner and the exercises are timed. Also, the problem of nonreaders is overcome through the oral directions. While the technology had been available prior to this use, it had not been widely used. The procedures for scoring subjective exercises have provided a model for accurate, objective scoring of these kinds of exercises. The scorers were put through a special training program for scoring essays, written materials, or performance exercises.

The results have been released in census-like fashion without interpretation. Clear, factual reports were made of the data so the reader would know what was assessed and what were the results. Very little interpretation of the results has been done for the reader. This was the intent of the National Assessment staff, for they did not conceive of their role as going beyond the reporting function.

The final two components of the model are the most controversial, for originally there was not a clear mandate for National Assessment to take a vigorous leadership role with respect to these two tasks. As time has gone by, greater activity by National Assessment in these areas has been urged by such groups as educators and parents, and National Assessment has become more active in the dissemination and utilization components.

In summary, it can be pointed out that the model reflects some important choices on the part of the staff of National Assessment. The decision was made to assess a broad range of objectives in each of the ten subject matter fields. Certainly, it would have been much easier and cheaper to have concentrated on a narrow set of objectives. Also, the effort was successfully made to include the higher cognitive levels in the assessment exercises and to deal with the affective domain. In subject areas like Citizenship and Social Studies, the usual standardized test concentrates on factual knowledge which is of a less controversial nature and easier to assess. National Assessment should be commended for their bolder, more comprehensive approach to the task which searches for more significant kinds of data.